John von Neumann Institute for Computing

From Simulation to Therapy: A Systems Biology Approach to Oncogene Detection

Avijit Ghosh, Dhruv Pant, Aparna Kumar, Rui Zou, Dave Miller

published in

NIC Workshop 2006,
From Computational Biophysics to Systems Biology,
Jan Meinke, Olav Zimmermann,
Sandipan Mohanty, Ulrich H.E. Hansmann (Editors)
John von Neumann Institute for Computing, Jülich,
NIC Series, Vol. **34**, ISBN-10: 3-9810843-0-6,
ISBN-13: 978-3-9810843-0-6, pp. 7-14 , 2006.

http://www.fz-juelich.de/nic-series/volume34

From Simulation to Therapy: A Systems Biology Approach to Oncogene Detection

Avijit Ghosh[1], **Dhruv Pant**[2], **Aparna Kumar**[3], **Rui Zou**[4], **and Dave Miller**[1]

[1] Department of Physics, Drexel University
3141 Chestnut St, Philadelphia, PA 19104

[2] Watson School of Biological Sciences, Cold Spring Harbor Laboratory
One Bungtown Road, Cold Spring Harbor, NY 11724

[3] Merrimack Pharmaceuticals
101 Binney St, Cambridge, MA 02142

[4] Department of Biomedical Engineering, Drexel University
3141 Chestnut St, Philadelphia, PA 19104

In silico models of signal transduction pathways have been successful both qualitatively as well as quantitatively in describing how complex protein networks control cell function. Moreover, the study of networks has been used to elucidate not only how these pathways control the complex regulation and response mechanism of cells, but also provide insight into how a breakdown in the biological circuitry can lead to particular disease states.

We have recently examined the circuitry within the MAPK signal transduction pathway to understand how changes within this canonical network may lead to malfunction, notably the rise of proto-oncogenic cells. In addition we have developed a new complementary technique that provides insight into which key players within the pathway are most likely to be most conducive to selective inhibition within this transformed line of cells. These tools have been made freely available to the public, as part of a software suite developed by our group, *Cellsim*[1]. I will give an overview on how *Cellsim* may be used to quantitate cell function and moreover malfunction.

1 Introduction

Computational Biophysics has always played a complementary role to the experimental biological sciences. The role of a computational biophysicist, as such, is not to develop tools that simply reassure the experimentalists that well-regarded experiments may, in fact, be duplicated *in silico*, but instead must also provide new predictive and quantitative tools that provide new insight into biological mechanism or function. New tools from the development of new experimentally designed united force fields such as UNRES[2] to new special purpose hardware techniques such as MDGRAPE[3] have given rise to new predictive mathematical and computational methods that can probe behavior of individual proteins on a femtosecond scale. It is in this sense that new tools such as systems biology have been developed to address the complementary issue of how these proteins can act *en masse* to dictate not only form, but intra-cellular behavior[4–8]. The scale is perhaps different (from angstrom to micrometer, from individual proteins to micromolar concentrations, from picosecond to millisecond and so forth), but the philosophy is the same. The method of systems biology is to take experimental data that is relatively simple to reproduce, and to use it to provide insight into phenomena that can not be easily elucidated by an existing set of experiments.

Systems biology uses these experimentally derived, computational techniques to elucidate both how cells have the ability to respond to external stimuli and how intra-cellular signaling circuitry that is essential for cell function is controlled. The mechanism by which this occurs depends fundamentally on the way in which cells use protein networks as the mechanism for translating extra-cellular signals into intra-cellular behavior. Hence, the complexity of signal transduction networks is based on the interplay between different aspects of the signaling process, any of which may change with subtle external or internal changes to the cell[9]. The focus of our group is on one particularly important signaling cascade: the MAPK signal transduction pathway.

2 The Canonical MAPK Signal Transduction Pathway

The canonical MAPK signaling cascade (Figure 1) is perhaps one of the most well studied signal cascades, both experimentally and computationally[10]. This central cascade is critical for governing cell growth and proliferation as well as actin cytoskeleton rearrangement[11–14]. Stimulation of the cascade activates many downstream effectors including PI3K[15], Bcl-2[16], and PKC,[17,18] among others[10]. The central cascade is activated via the following mechanism: The epidermal growth factor (EGF) signaling begins with the epidermal growth factor receptor (EGFR) and traverses a series of signaling proteins to the Ras protein. The Ras protein works in part by activating a series of kinases starting from Raf (a Mitogen Activated Protein Kinase Kinase Kinase (MAPKKK)) which activates the Mitogen Activated Protein ERK Kinase (MEK), This in turn activates the extracellular-signal regulation kinase (ERK), which subsequently translocates to the nucleus and stimulates a series of growth promoting transcription factors[19]. The pathways described represent a simplified description of the full process of cell signaling, as this cascade is but a single member of a complex set of parallel, interacting pathways[9].

2.1 Oncogenic Transformation of the MAPK Pxathway

The central member of the MAPK pathway, Ras, illustrates the importance of transformations within the MAPK signal transduction cascade. The Ras protein is a GTP-binding signaling protein, activating downstream effectors by binding to GTP, while inactive in the GDP bound form. The Ras oncogene was first found experimentally by its ability to induce tumor-like growth in fibroblasts[20–22]. The major oncogenic transformation involves specific mutations in Ras that prevent hydrolysis of GTP-bound Ras by GTPase Activating Proteins (GAP), leaving Ras continuously activated (and thus persistently signaling downstream effectors). These particular mutations have been implicated in approximately 30% of all human cancers, predominantly in lung, colon and pancreatic cancers[23]. This is likely an under-estimation of oncogenic transformation of Ras related cascades as mutations in other effectors downstream may cause a similar transformation in absence of mutations in the Ras gene itself. The transformation to tumor cells does not occur by a mutational event in Ras alone, but through a series of malignant transformations along or between several distinct pathway species[24].

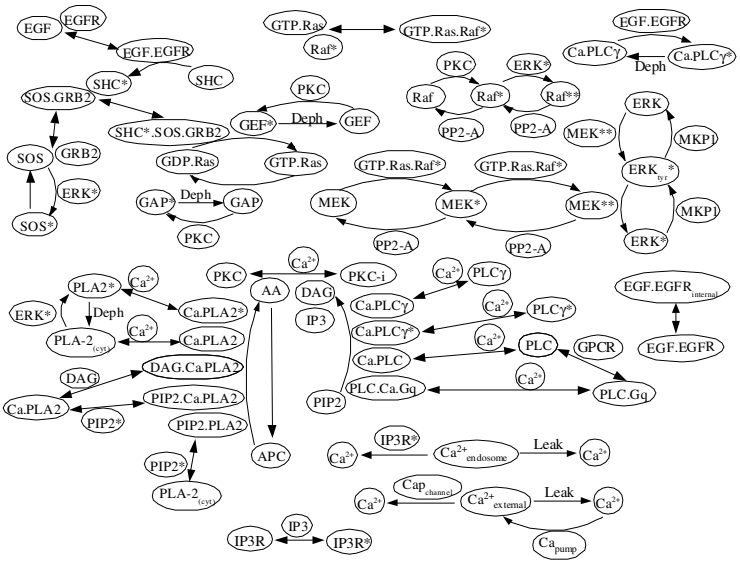

Figure 1. The canonical MAPK pathway

3 Methods

Systems biology methods developed by our group use either ordinary differential equations (ODEs) or partial differential equations (PDEs) to describe the overall temporal or spatio-temporal behavior of the protein network within (and between) compartments of a cell[7,8]. Enzymatic reactions and other chemical interactions are represented as simply a system of ODEs which couple to active and passive transport. Passive transport includes processes such as simple diffusive processes. Active transport includes explicit advective terms, modeling for instance transport along actin filaments and other ATP driven processes.

Elementary chemical reactions describe the enzymatic and non-enzymatic reactions within each compartment. These reactions may be written as:

$$\sum_i n_i R_i \underset{k_b}{\overset{k_f}{\rightleftharpoons}} \sum_j n_j P_j \tag{1}$$

where a set of reactant species R_i with stoichiometric coefficients n_i inter-converts into a set of product species P_j with stoichiometric coefficients n_j with rate constants k_f and k_b. As collisions that are greater than bimolecular are rare, the order of an elementary chemical reaction is not typically greater than 2. Characteristic of signaling pathways are enzymatic reactions such as phosphorylation or dephosphorylation events. These reactions may be expressed as a combination of a reversible and an irreversible chemical reaction as follows:

$$E + S \underset{k_2}{\overset{k_1}{\rightleftharpoons}} E.S \xrightarrow{k_3} E + S^* \tag{2}$$

where E represents an enzyme which catalyzes the substrate S. The intermediate species $E.S$ first forms reversibly with rate constants k_1 and k_2 followed by an irreversible catalytic step with rate constant k_3 which releases the activated substrate S^* and the enzyme for further catalysis.

These reactions lead to a set of ordinary differential equations such that one may express the time rate of change in concentration of all species as a system of unimolecular and bimolecular reactions such that:

$$\frac{d[C_i]}{dt} = \sum_j k_{ij}[C_j] + \sum_{l>m} k_{ilm}[C_l][C_m] + \sum_j T(C_i, C_j) \tag{3}$$

where k_{ij} is the rate constant for a unimolecular reaction involving species C_i and C_j at concentrations $[C_i]$ and $[C_j]$ respectively. If k_{ilm} is positive then k_{ilm} represents the rate constant of formation of species C_i from a bimolecular reaction between species C_l and C_m with concentrations $[C_l]$ and $[C_m]$. Conversely, if k_{ilm} is negative then k_{ilm} represents the rate constant of disassociation of species C_i into two species C_l and C_m. $T(C_i, C_j)$ represents a function governing the passive transport of a species C_i into a different compartment at which time it is labeled with a subscript j as C_j via passive channels.

3.1 Mutations

A mutation in a particular gene in a signaling pathway manifests itself in one of two ways. In the first case, the mutation may directly affect the *interaction* between two species. If two species A and B reversibly associate/disassociate with rate constants k_f/k_b:

$$A + B \underset{k_b}{\overset{k_f}{\rightleftharpoons}} A.B \tag{4}$$

then a mutation of this type will perturb k_f or k_b by some amount. For instance, lowering k_f by some amount represents a mutation that hinders the ability of species A or B to associate into $A.B$. We define a mutation of this type as an "interaction" mutation. This mutation may occur in either species A or B, as the effect is the same. A simple analysis using Arrhenius theory may be used to connect the free energy change from a mutation with the corresponding kinetic parameters k_f and k_b[6,5].

The rate of interconversion from reactant to product may be given as:

$$k_f = Ae^{-E_a/k_B T} \tag{5}$$

where A is a constant prefactor, E_a is the barrier energy of activation and k_B is Boltzmann's constant. Mutations in the enzyme may affect a transition rate by either increasing the barrier height or changing the free energy of the initial state of the system by some amount ΔE. The new mutated system may therefore be considered to be a perturbed system with a new barrier of height E'_a and a forward transition rate of k'_f (Figure 2). The ratio k'_f/k_f is :

$$k'_f/k_f = \frac{e^{-(E_a+\Delta E)/k_b T}}{e^{-E_a/k_b T}} \tag{6}$$

and simplifying

$$k'_f/k_f = e^{-\Delta E/k_b T} \tag{7}$$

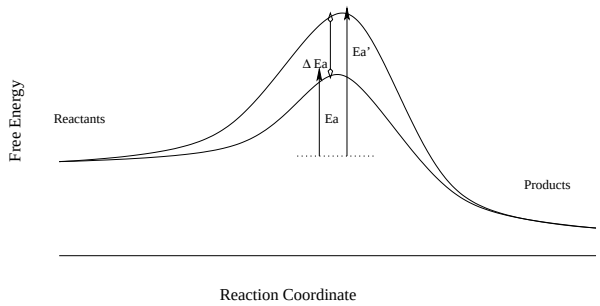

Figure 2. Graphical representation of the energy barriers involved in a reaction. An increase in the barrier height lowers the rate constant. The lower rate constant k'_f corresponds to the greater barrier height E'_a.

The right side of eq. (7) is a function of the *change* in the barrier height and not of the barrier height E_a itself.

Eq. 7 can be used to simulate the *effect* of any single mutation on the normal MAPK signal transduction pathway without having to explicitly dileanate the underlying cause. The key effect governing the transformation of the normal MAPK signal transduction pathway, is the ability to activate downstream ERK *without* EGF stimulation. The results from this analysis will be published in a forthcoming manuscript[25].

Moreover, such mutations may be rank ordered in terms of ΔE. The top 10 mutations (those that need the smallest ΔE increase to activate ERK) are shown in table 1.

Rank	k_{mod}	*Reaction*	*Dir*
1	0.75	X* + PP2A $\rightleftharpoons$ X*-PP2A	f
2	0.75	X*-PP2A $\rightarrow$ PP2A +X	f
3	0.6	Raf* +GTP-Ras $\rightleftharpoons$ Raf-GTP-Ras*	b
4	0.55	Raf* + PP2A $\rightleftharpoons$ Raf*-PP2A	f
5	0.55	GTP-Ras + GAP $\rightleftharpoons$ GTP-Ras-GAP	f
6	0.55	GTP-Ras-GAP $\rightarrow$ GAP +GDP-Ras	f
7	0.55	AA $\rightarrow$ APC	f
8	0.55	Ca-Ca$_{pump}$ $\rightarrow$ Ca$_{pump}$ +Ca$_{ext}$	f
9	0.5	X + Raf-GTP-Ras* $\rightleftharpoons$ X-Raf-GTP-Ras*	b
10	0.5	Ca + Ca$_{pump}$ $\rightleftharpoons$ Ca-Ca$_{pump}$	f

Table 1. Ranking of the 'interaction" mutations based on the highest value of k_{mod} below which ERK activation occurs. These reactions represent the "inhibiting" set of mutations. Reactions with X (or X*) or (g) represent group reactions.

4 Drug Targeting

With drug development costs now reaching 500 million dollars and more, development strategies represent a significant hurdle in bringing new therapeutics to the marketplace[26].

The experimental development cycle can be optimized in a manner that minimizes the number of false positives during costly clinical trials using systems biology techniques similar to those described previously in this manuscript to detect proto-oncogenes. Rather than modeling the effect of a mutation, modeling of an inhibitor may be performed with the addition of a single chemical reaction representing simple competitive binding between the substrate and the target protein:

$$I + S \underset{k_f}{\overset{k_b}{\rightleftharpoons}} I.S \tag{8}$$

The binding free energy can be calculated from the equilibrium constant k_{eq} of the reversible binding reaction. A particular inhibitor may bind to any substrate within the MAPK pathway. The efficacy of the inhibitor against a particular target is gauged by its ability to stop auto-activation of the pathway of the entire set of transformed cells described in the previous section. Furthermore, targets that successfully inhibit all transformed cell lines are further ranked by the minimum binding affinity and concentration needed to deactivate all cell lines.

5 Concluding Remarks

Computational biophysics has been successful in underscoring how quantitation and simulation can be used to address difficult problems of interest in biology, as evidenced by the many articles within this book. Systems biology continues this tradition with an emphasis on the *macroscopic* rather than the *microscopic*, focusing on not single molecules or proteins but rather how entire systems interact. As illustrated, systems biology can be used to not only quantitate how behavior is governed, but moreover how malfunctions in the signaling process can give rise to aberrant signaling processes. Finally, an analysis of these aberrant networks can be used to suggest novel treatment strategies.

References

1. A. Ghosh, D. Pant, A. Kumar, D. Miller, R. Zou, H. Lin, Y. Wang, and T. Hoppe. http:// silicon.physics.drexel.edu /cellsim, 2004. CellSIM Home Page.
2. A. Liwo, P. Arlukowicz, C. Czaplewski, S. Oldziej, J. Pillardy, and H.A. Scheraga. A method for optimizing potential-energy functions by hierarchical design of the potential-energy landscape: application to the unres force field. *PNAS*, 99:1937–42, 2002.
3. Y. Komeiji, H. Yokoyama, M. Uebayasi, M. Taiji, T. Fukushige, D. Sugimoto, R. Takata, A. Shimizu, and K. Itsukashi. A high performance system for molecular dynamics simulation of biomolecules using a special-purpose computer. *Pac Symp. Biocomput.*, pages 472–87, 1996.
4. R. Zou and A. Ghosh. Automated sensitivity analysis of stiff biochemical systems using a fourth-order adaptive step size rosenbrock integration method. *IEE Proc. Systems Biology*, 153:79–90, 2006.
5. D.K. Pant and A. Ghosh. A systems biology approach for the study of cumulative oncogenes, with applications to the mapk signal transduction pathway. *Biophys. Chem.*, 119:49–60, 2006.

6. D.K. Pant and A. Ghosh. Automated oncogene detection in complex protein networks, with applications to the mapk signal transduction pathway. *Biophys. Chem.*, 113:275–288, 2005.
7. A. Ghosh, D. Miller, R. Zou, B. Sokhansanj, and A. Kriete. *Spatio-Temporal Systems Biology*, pages 327–382. Elsevier Press, 2006.
8. A. Ghosh, D. Miller, R. Zou, H. Pais, B. Sokhansanj, and A. Kriete. Integrated spatio-temporal model of cell signaling. In *FOSBE 2005 Proceedings*. FOSBE, 2005.
9. U.S. Bhalla and R. Iyengar. Emergent properties of networks of biological signaling pathways. *Science*, 283:381–387, 1999.
10. T. Joneson and D. Bar-Sagi. Ras effectors and their role in mitogenesis and oncogenesis. *J. Mol. Med.*, 75:587–93, 1997.
11. C.J. Marshall. Ras effectors. *Curr. Opin. Cell. Biol.*, 8:197–204, 1996.
12. A.F. Chambers and A.B. Tuck. Ras-responsive genes and tumor metastasis. *Crit Rev Oncog*, 4:95–114, 1993.
13. A.B. Vojtek and C.J. Der. Increasing complexity of the ras signaling pathway. *J. Biol. Chem.*, 273:19925–19928, 1998.
14. D.T. Denhardt. Signal-transducing protein phosphorylation cascades mediated by ras/rho proteins in the mammalian cell: the potential for multiplex signalling. *Biochem J*, 318:729–47, 1996.
15. B.M. Marte and J. Downward. Pkb/akt: connecting phosphoinositide 3-kinase to cell survival and beyond. *Trends Biochem Sci*, 22:355–8, 1997.
16. T. Tsuchida, H. Kijima, S. Hori, Y. Oshika, T. Tokunaga, K. Kawai, H. Yamazaki, Y. Ueyama, K.J. Scanlon, N. Tamaoki, and M. Nakamura. Adenovirus-mediated anti-k-ras ribozyme induces apoptosis and growth suppression of human pancreatic carcinoma. *Cancer Gene Ther*, 7:373–83, 2000.
17. F. Uberall, K. Hellbert, S. Kampfer, K. Maly, A. Villunger, M. Spitaler, J. Mwanjewe, G. Baier-Bitterlich, G. Baier, and H. H. Grunicke. Evidence that atypical protein kinase c-lambda and atypical protein kinase c-zeta participate in ras-mediated reorganization of the f-actin cytoskeleton. *J. Cell. Biol.*, 144:413–425, 1999.
18. S. Kampfer, F. Uberall, S. Giselbrecht, K. Hellbert, G. Baier, and H. H. Grunicke. Characterization of pkc isozyme specific functions in cellular signaling. *Adv Enzyme Regul.*, 38:35–48, 1998.
19. S. J. Mansour, W. T. Matten, A. S. Hermann, J. M. Candia, S. Rong, K. Fukasawa, G. F. Vande Woude, and N. G. Ahn. Transformation of mammalian cells by constitutively active map kinase kinase. *Science*, 265:966–70, 1994.
20. L. F. Parada, C. J. Tabin, C. Shih, and R. A. Weinberg. Human ej bladder carcinoma oncogene is homologue of harvey sarcoma virus ras gene. *Nature*, 297:474–8, 1982.
21. C. J. Tabin, S. M. Bradley, C. I. Bargmann, R. A. Weinberg, A. G. Papageorge, E. M. Scolnick, R. Dhar, D. R. Lowy, and E. H. Chang. Mechanism of activation of a human oncogene. *Nature*, 300:143–9, 1982.
22. R. W. Ellis, D. DeFeo, J. M. Maryak, H. A. Young, T. Y. Shih, E. H. Chang, D. R. Lowy, and E. M. Scolnick. Dual evolutionary origin for the rat genetic sequences of harvey murine sarcoma virus. *J Virol*, 36:408–20, 1980.
23. J.L. Bos. Ras oncogenes in human cancer: a review. *Cancer Res.*, 49:4682–4689, 1989.
24. R.A. Weinberg. Oncogenes, antioncogenes, and the molecular bases of multistep

carcinogenesis. *Cancer Res*, 49:3713–21, 1989.
25. A. Kumar, R. Zou, and A. Ghosh. A systems biology approach for lead drug target prediction with applications to the mapk signal transduction pathway. In Prep.
26. R.G. Frank. New estimates of drug development costs. *J. Health Econ.*, 22:325–30, 2003.